DOCUMENT RESUME

ED 230 582 TM 830 332

AUTHOR Tollefson, Nona; Tripp, Alice
TITLE The Effect of Item Format on Item Difficulty and Item Discrimination.
PUB DATE Apr 83
NOTE 19p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, Montreal, Quebec, April 11-15, 1983).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Variance; *Difficulty Level; Graduate Students; Higher Education; Item Banks; *Multiple Choice Tests; *Responses; Statistics; *Test Format; Testing; *Test Items
IDENTIFIERS *Complex Response Alternatives; *Item Discrimination (Tests)

ABSTRACT

This study compared the item difficulty and item discrimination of three multiple choice item formats. The multiple choice formats studied were: a complex alternative (none of the above) as the correct answer; a complex alternative as a foil, and the one-correct answer format. One hundred four graduate students were randomly assigned to complete one of three forms of a multiple choice test. Statistical concepts were the content of the test. Analysis of variance showed significant mean differences in the item discrimination indices for the three item formats. Items having "none of the above" as the correct answer had a significantly higher mean discrimination index than the other two item formats. There was no significant difference in the mean item difficulties for the three item formats. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED230582

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

N. Tollefson

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Effect of Item Format on Item Difficulty and Item Discrimination

Nona Tollefson
Associate Professor
Educational Psychology and Research
The University of Kansas

Alice Tripp
Assistant Professor
School of Nursing
The University of Arizona

This paper was presented at the annual meeting of the American Educational Research Association in Montreal, April, 1983.

TM 830 332

The Effect of Item Format on Item Difficulty and Item Discrimination

ABSTRACT

This study compared the item difficulty and item discrimination of 3 multiple choice item formats. The multiple choice formats studied were: a complex alternative (none of the above) as the correct answer; a complex alternative as a foil, and the one-correct answer format. One hundred four graduate students were randomly assigned to complete one of three forms of a multiple choice test. Statistical concepts were the content of the test. Analysis of variance showed significant mean differences in the item discrimination indices for the three item formats. Items having "none of the above" as the correct answer had a significantly higher mean discrimination index than the other two item formats. There was no significant difference in the mean item difficulties for the three item formats.

Controversy exists in the test construction field about the use of complex response alternatives in multiple choice items. Many measurement textbooks recommend the use of the "one right answer" multiple choice format (Benion, 1977; Brown, 1976; Kryspin & Feldhusen, 1974; Mehren & Lehmann, 1978; Scannell & Tracy, 1975). However, other measurement experts recommend the use of complex response alternatives (Bloom, Madaus, & Hastings, 1981; Ebel, 1979)

At the present time, the research literature describing the properties (item difficulty and item discrimination) of complex response alternatives is contradictory, and textbook writers give anecdotal reasons rather than empirical data for their recommendations to use or not to use complex response alternatives. A brief review of the literature on the use of complex response alternatives will illustrate the contradictory results characterizing research on complex response alternatives.

Wesman and Bennet (1946) reported no effect on item difficulty associated with the "none of the above" alternative. Boynton (1950), Rimland (1960), and Williamson and Hopkins (1967) all reported that the

use of the "none of the above" alternative increased item difficulty. Researchers studying the "none of the above" response option have all used arithmetic items as their test content. However, no information about the validity or reliability of the tests was included.

Studies of the effect on item difficulty and discrimination of complex alternatives of the "all of the above" and "A and B are correct" form have also yielded contradictory results. Hughes and Trible (1965) compared the single correct answer to the "both 1 and 2 are correct," the "all of the above" and "none of the above" alternatives. These researchers found the mean score for the test form including complex alternatives to be significantly lower than the mean score for the test form including items having only one correct answer.

Rossie, McCrady and Paoline (1978) compared the item discrimination of one correct response items and matched sets of items of the "A and B, but not C" form. These authors found no significant difference in the average discrimination index of the two forms of the test.

Weiten (1982) compared single and double option multiple choice items and found that double multiple choice items were more difficult and less discriminating than single option items. The set of double option multiple choice items also had lower internal consistency reliability estimates than the set of single option multiple choice items.

Purpose of the Study

This study compared the item difficulty and item discrimination of sets of multiple choice items having a single response as the correct answer and a complex alternative (none of the above) as the correct answer. Two studies were conducted. The first study compared experimental sets of multiple choice items having single and complex

options as the correct response. The second study compared non-experimental sets of multiple choice items having single and complex response options as the correct answer.

Study I

Subjects

Subjects for this study were graduate students in Education enrolled in a beginning statistics class. Four sections (n=104) of the statistics course participated in the study. The sections were taught by 2 instructors; each instructor taught 2 sections. The sample included 68 females and 36 males.

Instrument

The test instrument was the multiple choice section of the final examination in the statistics course. Thirty-six multiple choice items were randomly selected from an item pool of 250 items. The item pool had been developed by four professors teaching the basic statistics courses. All items had been field-tested and the item difficulty and item discrimination indices were known.

From the set of 36 multiple choice items, 10 items were selected as the experimental set of items. Seven of the items had verbal responses as the options. Three of the items required numerical calculations and had numerical responses as the options. Three forms of each of the 10 items were developed. One form of the item used the "one correct answer" format, the second form of the item used the "one correct answer" format with none-of-the above as a foil. The third form of the item had "none of the above" as the correct answer.

The 10 items were randomly assigned to one of three forms of the final examination. Form A included 3 items for which "none of the above" was the correct answer. Forms B and C included 4 items having "none of the above" as the correct response. Forms A and C each included 4 items where "none of the above" was a foil. Form B included two items using "none of the above" as a foil. The "one correct answer" format was used for 3 of the 10 items on Form A. Forms B and C each included 4 items using this format.

In addition a set four items using "the correct answer is not given" option was included on each form of test. This option is commonly found in multiple choice items measuring quantitative skills. The same set of 4 items was included on all forms of the test.

Test forms were randomly assigned to students at the time of the final examination. Form A was taken by 38 students. This form of the test had a mean of 25.7, a standard deviation of 3.37, and an internal consistency reliability of .73. Form B was completed by 34 students. The mean for Form B was 26.5, the standard deviation was 3.56 and the internal consistency reliability was .58. Form C completed by 32 students, had a mean of 27.9, a standard deviation of 3.36, and an internal consistency reliability of .55.

Data Analysis

Item difficulty and item discrimination indices were computed for all items on the final examination. Items were then divided into four items sets: 1) the set of 10 experimental items having "none of the above" as the correct answer (E-1), 2) the set of 10 experimental items having "none of the above" as a foil (E-2), 3) the set of 10 experimental items using the "one correct answer" format (E-3), and 4) the set of 4 items using "the correct answer is not given" format (C-4).

One way analysis of variance was used to test for over-all significant differences in the mean item difficulty and discrimination indexes for the 3 experimental sets of items. Post-hoc t-tests were used to identify group differences in cases where the omnibus F test was significant.

Results

Means and standard deviations were computed for the item difficulty and the item discrimination indexes for each of the four items sets. Table 1 reports the means and standard deviations. Data in Table 1 show that the item sets where "none of the above" was the correct answer (E-1) and where "no correct answer given" was the correct answer (C-1) were the most difficult and most discriminatory sets of items.

Insert Table 1 Here

One-way analysis of variance was used to test for significant mean differences in item difficulty and item discrimination for the three sets of experimental items. The analysis of variance showed no significant difference in mean item difficulty values for the 3 experimental sets of items ($F(2,27) = 1.18$, $p. > .05$), but significant differences in the mean item discrimination values were found ($F(2,27) = 4.40$, $p. < .05$). t-tests showed the mean discrimination index for the first experimental set of items (none of the above as the correct answer) was significantly higher than for second ($t(18) = 3.81$ $p < .01$) and third ($t(18 = 2.68$ $p < .05$) experimental sets of items. There was no statistically significant difference between the mean item discrimination indexes for the second and third experimental sets of items.

The multiple choice test included a total of 14 items where a complex alternative was the correct answer (10 items where "none of the above" was the correct answer and 4 items where "no correct answer given" was the correct answer). Seven of the 14 items presented verbal concepts and seven presented quantitative concepts. Table 2 shows the mean item difficulty values for the three sets of experimental items having verbal and quantitative response options. Table 3 gives the mean item discrimination indexes for these experimental sets of data. These data showed the set of verbal items to be more difficult than the set of quantitative items across all three item formats. A t-test for difference in the mean item difficulty indexes for verbal and quantitative sets of items having a complex alternative as the correct response yielded non-significant results. However, the item discrimination data showed the complex alternative format had a significantly higher mean discrimination index ($t- = 2.30$, $p < .05$) for the verbal than the quantitative set of items.

Insert Tables 2 and 3 here

Study II

Subjects

The second study compared the item characteristics of a non-experimental set of multiple choice items administered as a final examination in a first year graduate statistics course. Forty-four graduate students completed the examination.

Instrument

The final examination consisted of 42 multiple choice items. Of the 42 items, "None of the above" was the correct response for 7 items; "None of the above" was a foil for 10 of the items; "a and b are correct" was a foil in 5 of the items, the remaining 20 items had one correct answer format and single options as foils. The final examination had a mean of 28.9 and a standard deviation of 8.64. The Kuder-Richardson internal consistency reliability was .91.

Data Analysis and Results

Item means and item standard deviations were computed for the item difficulty and item discrimination indexs for the sets of items having one correct answer and a complex alternative as the correct answer. Table 4 reports the means and standard deviations for item difficulty and item discrimination indexes for the single option and multiple option items. A t-test for independent groups was used to test for significant mean difference in item difficulty for these two sets of items. The t-test yielded a significant difference in the mean item difficulty for the two sets of items ($t(40) = 2.17$ $p < .05$). Items with one correct answer were significantly easier than items with complex alternatives as the correct answer.

Insert Table 4 here

The point-biserial correlations used to measure item discrimination were transformed to Fisher's z' values to normalize the distribution of item discrimination values. A t-test was used to test the hypothesis of no difference in mean z´ values for the discrimination indices for

one correct answer and complex alternative multiple choice items. This analysis yielded non-significant results ($t(40) = .43$ ns).

Inspection of the mean item difficulty and discrimination indices for the different types of multiple option items showed no significant differences in mean item difficulty or discrimination for the sets of multiple option items. Table 5 reports the mean item characteristics for the multiple option multiple choice items

---

Insert Table 5 here

DISCUSSION

This study investigated the item characteristics of multiple choice items using complex alternatives and one-correct answer formats in an experimental and a non-experimental setting. In the experimental study (Study I), the use of the complex alternative as the correct answer and as a foil were compared to each other and to the one-correct answer format. Items with the complex alternative as the correct response were consistently more difficult than the other two item formats. However, the differences in item difficulty while consistent across six comparisons were not large enough to be statistically significant.

The non-experimental study also found that items having a complex alternative as the correct answer were more difficult than the set of items having one correct answer. In the second study, the differences in means were statistically significant.

In Study I, the complex alternative when used as the right answer showed significantly greater mean item discrimination than the other two item formats. No significant difference in mean item discrimination

values was observed in Study II. It has been argued that complex alternatives like the "none of the above" response require more knowledge of the content and make finer discrimination among test-takers. This study provided limited empirical support for this argument.

The use of "none of the above" or "no correct answer given" as the correct answer is common in quantitative tests. This study found items of this type were more difficult and discriminated better than items giving the correct answer. It has been argued that knowing you have solved a quantitative problem correctly and that the correct answer is not given requires a higher level of knowledge and/or comprehension than trying to solve a problem several ways in the hope of "finding one of the listed response options." This study offers limited support for this point of view.

This study investigated the use of complex alternatives as the correct answer in experimental and non-experimental test settings. Both studies showed items with complex alternatives to be more difficult than items with a one correct answer format. The results of the analysis of the item discrimination data were inconclusive. The analysis of the experimental set of items showed items with complex alternatives as the correct answer to be more discriminatory than items using the one correct answer format. The non-experimental study did not support this finding.

The study, while limited in scope, suggests the need for further study of the properties of multiple choice items using complex alternatives. Understanding the factors that influence students' responses to multiple choice items is important both theoretically and practically. Multiple choice items are widely used in achievement testing. Any information which increases test constructors' knowledge of the

strengths and weaknesses of various forms of multiple choice items will add to test developers' skill in measuring student achievement.

References

Bennion, D. H. Assessing Student Learning; Behavioral Objectives and Evaluation Procedures for the Affective, Cognitive and Psychomotor Domains. Dubuque, Iowa: Kendall-Hunt Publishing Company, 1977.

Bloom, B. S., Madaus, G. F. & Hastings, J. T. Evaluation to Improve Learning. New York: McGraw-Hill, 1981.

Boynton, M. Inclusion of 'none of these' makes spelling items more difficult. Educational and Psychological Measurement, 1950, 10, 431-432.

Brown, F. A. Principles of Educational and Psychological Testing (2nd ed.). New York: Holt, Rinehart & Winston, 1976.

Ebel, R. L. Essentials of Educational Measurement (3rd ed). Englewood Cliffs, New Jersey: Prentice Hall, Inc., 1979.

Kryspin, W. J. and Feldhusen, J. F. Developing classroom Tests: A Guide for Writing and Evaluating Test Items. Minneapolis, Minnesota: Burgess Publishing Co., 1974.

Mehren, W. A., and Lehmann, I. J. Measurement and evaluation in education and psychology (2nd ed.). New York: Holt, Rinehart & Winston, 1978.

Rimland, B. The effect of varying time limits and use of 'right answer not given' in experiment forms of the U.S. Navy Arithmetic Test. Educational and Psychological Measurement, 1960, 20, 533-539.

Scannell, D. P. and Tracy, D. B. Testing and Measurement in the Classroom. Boston: Houghton Mifflin Company, 1975.

Weiten, W. Relative effectiveness of single and double multiple-choice questions in educational measurement. Journal of Educational Research, 1982, 5, 46-50.

Wesman, A. G. and Bennett, G. K. The use of 'none of these' as an option in test construction. Journal of Educational Psychology, 1946, 27, 541-549.

Williamson, M. L. and Hopkins, K. D. The use of 'none of these' versus homogeneous alternatives on multiple choice tests: experimental reliability and validity comparisons. Journal of Educational Measurements, 1967, 4, 53-57.

Table 1

Means and Standard Deviations for Item Difficulty and Item Discrimination for Five Sets of Item Using Different Correct Answer Formats

Item Set		Item Difficulty $\bar{X}$	Item Difficulty s	Item Discrimination $\bar{X}$	Item Discrimination s
E-1	None of the above is the correct response	.69	.21	.36	.14
E-2	None of the above is a foil	.75	.20	.13	.13
E-3	One correct answer	.82	.16	.18	.16
C-1	No correct answer give (computation problems)	.68	.13	.27	.16

Table 2

Mean Item Difficulty for Three Multiple Choice Items Formats having Verbal and Quantitative Response Options

Type of Item Format	Type of Response			
	Quantitative		Verbal	
	$\bar{X}$	s	$\bar{X}$	s
Complex Alternative as the correct answer	.69	.17	.62	.21
Complex Alternative as a foil	.80	.14	.69	.21
One-correct answer	.80	.15	.78	.17

Table 3

Mean Item Discrimination Indexes for Three Multiple Choice Item Formats Having Verbal and Quantitative Response Options

Type of Item Format	Type of Response			
	Quantitative		Verbal	
	$\bar{X}$	s	$\bar{X}$	s
Complex Alternative as the correct response	.28	.13	.41	.13
Complex Alternative as a foil	.16	.14	.14	.12
One-correct answer	.26	.19	.19	.16

Table 4

Means and standard deviations for Item Difficulty and Item Discrimination for Single Option and Multiple Option Multiple Choice Items

Type of Item	Item Difficulty		Item Discrimination	
	$\bar{X}$	S	$\bar{X}$	S
One correct answer	.72	.18	.50	.14
Complex alternative as the correct answer	.59	.22	.47	.18

Table 5

Means and Standard Deviations for Item Difficulty and Item Discriminating for the Sub-sets of Multiple Option Multiple Choice Items

Type of Item	Item Difficulty $\bar{X}$	Item Difficulty S	Item Discrimination $\bar{X}$	Item Discrimination S
"None of the Above" as the Correct answer	.58	.21	.46	.12
"None of the Above" as a foil	.59	.26	.44	.23
"a & b are correct" as a foil	.61	.16	.53	.14